Teacher Characteristics and Their Effects on Student Test Scores: A Systematic Review.

Coenen, Johan; Cornelisz, Ilja; Groot, Wim; Maassen van den Brink, Henriette; Van Klaveren, Chris

published in
Journal of Economic Surveys
2018

DOI (link to publisher)
10.1111/joes.12210

document version
Publisher's PDF, also known as Version of record

JOURNAL OF ECONOMIC SURVEYS

doi: 10.1111/joes.12210

TEACHER CHARACTERISTICS AND THEIR EFFECTS ON STUDENT TEST SCORES: A SYSTEMATIC REVIEW

Johan Coenen*

Etil Research Group
the Netherlands

Ilja Cornelisz

Amsterdam Center for Learning Analytics
VU University Amsterdam
the Netherlands

Wim Groot and Henriette Maassen van den Brink

TIER
Maastricht University
the Netherlands

Chris Van Klaveren

Amsterdam Center for Learning Analytics
VU University Amsterdam
the Netherlands

Abstract. It has become widely accepted that teachers are important in facilitating student learning. Hundreds of empirical studies have tried to explain differences in student performance by evaluating the impact of particular teacher characteristics. Yet, this topic has not been the subject of a systematic review for more than 10 years, even though most of the empirical evidence has emerged over the past decade. This study provides an up-to-date review, drawing on empirical findings from several countries and distinguishing between acquired and sociodemographic teacher characteristics. This review confirms the existing consensus that subject-related degrees and knowledge, and not general teacher certifications, are positively related to student performance and particularly so for Master's degrees in math and science. A new insight is that recent findings point out that teacher experience continues to contribute to student test scores throughout a teacher's career, instead of merely the first few years. An important future research avenue would be to examine which mechanisms can explain these teacher characteristic effects.

Keywords. Student performance; Teacher characteristics; Teacher quality; Test scores

*Corresponding author contact email: jb.coenen@gmail.com; Tel: +31615317482.

Journal of Economic Surveys (2018) Vol. 32, No. 3, pp. 848–877

1. Introduction

Teachers are generally considered to be major contributors to student learning. Hundreds of empirical studies have examined the importance of teachers in explaining observed differences in student performance (Hanushek, 2011). Wayne and Youngs (2003) review the body of empirical evidence available up until 2001 and find that students generally learn more from teachers with higher test scores and higher college ratings. The evidence with respect to the impacts of degrees, coursework and certification is inconclusive. A notable exception in this respect is mathematics: high school students perform better when a teacher is certified in, and/or has a degree or completed coursework related to, mathematics. The intricate relationship between teachers and student performance, as measured by their test scores, has not been the subject of a systematic literature review for over 10 years. Yet, most of the available empirical evidence has emerged over the past decade (Figure 1). This more recent work has the potential to deliver important new insights. Therefore, this study provides an up-to-date systematic review of the evidence on how observed teacher characteristics relate to student test scores.

The empirical literature on teacher characteristics, generally distinguishes between two different dimensions of teacher aspects, defined here as acquired and sociodemographic teacher characteristics. Acquired teacher characteristics refer to education- and experience-related characteristics (advanced degrees, college quality, teaching certificates, teacher test scores and years of teaching experience) which, for individual teachers, can vary over time. Studying these is of great importance as these characteristics can thus be influenced directly through education policy instruments. The empirical literature on sociodemographic (or background) teacher characteristics generally examine the effects of fixed aspects, most notably teacher gender and/or ethnicity. Even though education policy cannot directly impact on these characteristics at the level of the individual teacher, empirical findings indicate that their contributions in explaining differences in performance are substantial, particularly in light of an increasingly diversified student population. A thorough understanding of the mechanisms underlying these sociodemographic teacher effects (e.g. nature and quality of student-teacher interactions) may well

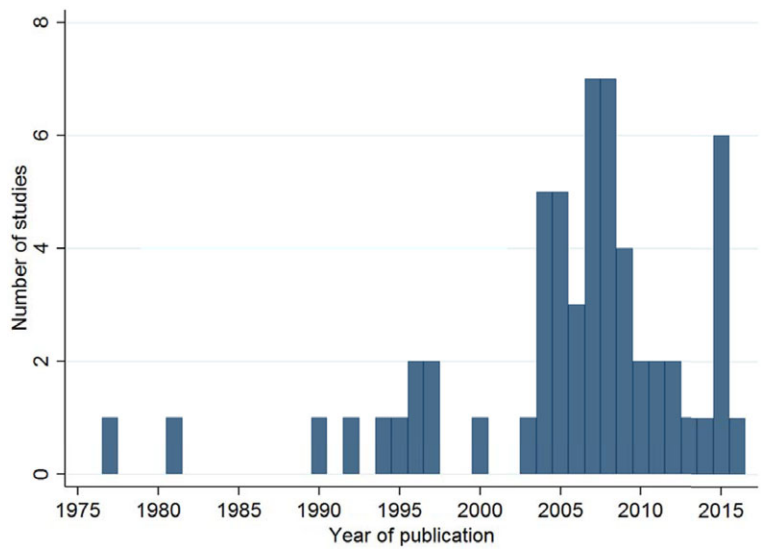

Figure 1. Number of studies on the influence of teacher characteristics on student test scores, per year. [Colour figure can be viewed at wileyonlinelibrary.com]

prove to be pivotal in understanding and improving teacher quality. As such, this review focuses on both these sociodemographic and the aforementioned acquired teacher characteristics.

The empirical literature, and therefore this review, focuses primarily on teacher effects on (standardized) student test scores. This is, at least partially, because data on test scores is often readily available. In addition, (standardized) test scores are generally considered to be a good approximation of actual learning and performance. Furthermore, standardized test scores have become increasingly central to policies regarding incentives and accountability in many educational contexts. As such, test scores seems an appropriate focus to review teacher effects. At the same time, it is important to recognize that teacher characteristics also influence other important student outcomes, such as absenteeism, noncognitive skills, and labour market outcomes. In this respect, the focus on test scores is a limitation. Future review studies could therefore contribute by focusing on the relationship between teacher characteristics and other educational outcomes.

This review applies a systematic search procedure and imposes inclusion criteria pertaining to relevance, study design and publication characteristics. The overall objective is to generate a comprehensive review of evidence, based only on high-quality studies. The multistep search strategy yielded 96 empirical studies. Imposing the inclusion criteria for high-quality evidence leave the results of 58 of these studies to be considered in this review. Of these 58 studies, 14 studies are not published in a peer-reviewed journal or thesis. The findings of these two sets of studies are contrasted to examine the potential implications of publication bias.

This review proceeds as follows. Section 2 describes the search method and inclusion criteria determining which empirical studies are considered in this review. Section 3 summarizes the empirical findings related to acquired teacher characteristics. Section 4 summarizes the empirical findings related to teacher gender and ethnicity. Section 5 concludes.

2. Method

2.1 *Search Method*

This literature review applies a four-step search method. In the first step, electronic databases were searched for literature on teacher characteristics and student achievement. The databases consulted are: Sage Journals Online, Jstor, PsychLit, EconLit and Google Scholar. At first, relatively generic search terms were chosen, such as 'teacher effectiveness', 'teacher quality', 'student performance', 'student test score' and 'teacher characteristics student achievement'. The studies collected using these more general terms can be classified into six categories: teacher education, teacher certification, teacher test scores and ability, teacher experience, teacher gender and teacher ethnicity. In step two, these specific categories were applied as search terms, such as 'teacher education student achievement', 'teacher certificate student achievement', 'teacher experience student achievement'. In step three, additional studies are collected by applying the 'snowball principle' (i.e. examining the reference list of each study and include all studies that focus on teacher characteristics and student test scores that were not yet included). The 'snowball principle' is reiterated as long as new studies are found. In the fourth and final step, the personal websites from authors of selected studies are consulted, as to gather additional studies relevant to effects of teacher characteristics on student test scores. Applying this four-step search method resulted in a total of 96 studies, which form the basis for this literature review.

2.2 *Inclusion Criteria*

Inclusion criteria are formulated in order to take into account the quality of the evidence provided by each of the 96 studies generated in the multistep search strategy. Studies are only considered in this review if the following six inclusion criteria are met:

1. Studies must use data with information on teacher characteristics and student test scores of standardized tests.
2. Studies must account for students' prior achievement.
3. Studies must account for socioeconomic status.
4. Studies must use a (quasi-)experimental or panel data design.
5. Studies must have same focus as review.
6. Studies must be published in a peer-reviewed journal or in a thesis.

The first three inclusion criteria are identical to those imposed by Wayne and Youngs (2003). The first criterion ensures that studies use data in which teacher characteristics can be linked to standardized student test scores. The second and third criterion recognize that student achievement levels are the result of cumulative process and thus influenced by previous teachers and performance, and that potential teacher contributions depend to a large extent on the socioeconomic background characteristics of students. Therefore, it is important that study designs control for differences in initial student achievement and socioeconomic background variables.

The next three inclusion criteria extend on those used by Wayne and Youngs (2003), with the objective to include only high-quality empirical evidence regarding student test scores. The fourth inclusion criterion (*design*) imposes that only evidence based on (quasi-)experimental or panel data studies is considered for this review.[1] The general consensus is that unbiased treatment effects can only be obtained by conducting (quasi-)experimental studies (see, among others, Angrist and Pischke, 2008; Blundell and Dias, 2009). However, there are virtually no studies on the effects of teacher characteristics on student performance in which both students *and* teachers are both randomly assigned to classes. As a result, panel data studies satisfying the first three inclusion criteria provide the best available evidence on how teacher characteristics affect student achievement.

One concern, when considering the results of panel data studies, is that these findings may give a distorted image of how teacher characteristics affect student performance when students and teachers are indeed not randomly matched. However, Chetty et al. (2014) showed that most of the bias that results from such nonrandom sorting of students to teachers, can be accounted for by controlling for prior student test scores. For this review, this implies that panel data results are informative, and taken into account, if the first three inclusion criteria are met.

Four types of panel data studies are distinguished in this review:

Type 0	Panel data studies which make use of Project STAR data
Type I	Panel data studies in which students are randomly assigned to classes
Type II	Panel data studies that consider the potential effects of unobserved heterogeneity
Type III	Other panel data studies

Type 0 studies make use of data from Project STAR, and stand out from the other panel data studies in that students and teachers were indeed randomly assigned to classes. This ensures that the estimated teacher effects on student performance cannot be driven by unobserved selective assignment of teachers and students to classes. However, a drawback is that the Project STAR experiment was deliberately conducted to examine the effect of class size reduction on student performance. It is possible that class size reduction effects interfere with the teacher effects on student performance (Dee, 2004; Nye et al., 2004).

Type I studies use data in which students are randomly assigned to classes, such that estimated teacher effects cannot be the result of selective student assignment to classes *within* schools. A limitation is that the selective sorting of teachers and students across schools is not accounted for. Type II studies recognize

that unobserved heterogeneity may bias the estimation results and examine to what extent this potentially affects the estimated effects. Type III studies include a rich set of control variables and assume that this is enough to account for the potential bias caused by unobserved heterogeneity.

Type I studies have an advantage over Type II and Type III studies, in that these studies better control for the potential selective assignment of students to classes within schools. Type II studies have an advantage over Type III studies, as these studies recognize that unobserved heterogeneity may potentially impose a bias on the estimated effect of teacher characteristics and perform robustness checks to address this potential bias.

Teachers may have been selectively assigned to classes/schools in Type I, II and Type III studies, which can bias the reported effects of teacher characteristics on student performance. All these studies control for a rich set of teacher, student and school characteristics and it must be assumed that this is sufficient to also control for the selective distribution of teachers across classes and schools. This identifying assumption cannot be verified empirically.

The fifth inclusion criterion (*focus*) ensures that studies are only included if the focus is appropriate for this review. Several panel studies satisfy the first two inclusion criteria, present estimates on which teacher characteristics are associated with student achievement gains, but do not explicitly examine how teacher characteristics are related to these gains. Instead, these teacher characteristics are merely used as control variables (see e.g. Schwerdt and Wuppermann, 2011; Van Klaveren, 2011). Therefore, given the lack of an appropriate focus and corresponding robustness and sensitivity checks, these studies are not included in this review.

The sixth inclusion criterion (*publication*) is that studies must be accepted for publication in international and peer-reviewed journals or that they are published as a chapter in a peer-reviewed thesis. This criterion is somewhat controversial though, because a well-known issue is that studies with positive and significant findings are more likely to be published (the so-called publication bias). It implies that due to this sixth inclusion criterion, this review may present findings that are positively biased. At the same time, this inclusion criterion reduces the probability that empirical results are subject to analytical or data problems, because peer-reviewed publications are far more likely to recognize and address such analytical or data deficiencies (Van Klaveren and De Wolf, 2015). To address the problem of publication bias, this review does consider the results of unpublished studies which satisfy the other five inclusion criteria. This review then examines whether the findings of these studies are indeed different from results which have been published in an international peer-reviewed journal.

In the end, a total of 40 of the 98 studies collected did not satisfy the first five inclusion criteria formulated above. This leaves 58 studies to be included in this review, of which 14 studies are not published in a peer-reviewed journal or thesis. The findings of these studies will be contrasted with findings which have been published, as to get a sense of the implications of potential publication bias.

2.3 *Description of the Selected Literature*

Figure 1 shows the number of studies included in this review by publication year. Most studies that satisfy the inclusion criteria imposed in this review appeared after 2003, the year of the aforementioned systematic review by Wayne and Youngs (2003). This illustrates the relevance of providing an update.

Table 1 contains the studies published in peer-reviewed journals, while Table 2 describes studies that are not (yet) published in such journals. Both tables display information regarding the author(s), the research design used, the observation years of the data used, the student test score measure(s), the grades considered, and (if available) the number of students and teachers used in the empirical analysis.

By far, most studies focus on the US context. In fact, no more than 7 of the 44 published studies do not use US data. Similarly, only 2 of the 14 unpublished studies are not based (exclusively) on US data. Most results are based on panel data designs (PD) and only a few studies deploy an experimental

Table 1. Description of Studies Included in Literature Review: Published in Peer-Reviewed Journals

Study	Method	Observation years	Measure	Grades	Students	Teachers	Country
Aaronson et al. (2007)	PD	1996–1998	Gains	9	52,957	589	USA
Coenen and Van Klaveren (2016)	PD	2010–2011	Level, M	3–5	2586	174	The Netherlands
Bosshardt and Watts (1990)	PD	1987	Level, pre	12	2383	94	USA
Boyd et al. (2006)	PD	1998–2003	Gains	3–8	960,970	3766	USA
Boyd et al. (2008)	PD	2000 and 2005	Gains	3–8	578,630		USA
Carrell and West (2010)	EX2	2000–2007	Gains	HE	10,534	91	USA
Clotfelter et al. (2006)	PD	2000–2001	Level, pre	5	60,791	3233	USA
Clotfelter et al. (2007)	PD	1994–2003	Both	3–5	180k		USA
Clotfelter et al. (2010)	PD	1999–2002	WSBS	10	137,597		USA
Croninger et al. (2007)	PD	1998 and 2000	Gains	1	5167	1342	USA
Darling-Hammond (2005)	PD	1996–2001	Gains	4–5	132,071	4408	USA
Dee (2004)	EX1	1985–1989	Level	K-3	8310		USA
Dee (2007)	PD	1988	WSBS	8	21,324		USA
Dee and Cohodes (2008)	PD	1988	WSBS	8	16,901		USA
Ehrenberg et al. (1995)	PD	1988 and 1990	Gains	10	3029		USA
Goldhaber and Anthony (2007)	PD	1996–1998	Gains	3–5	611,517	32,448	USA
Goldhaber and Brewer (1997)	PD	1988 and 1990	Gains	10	5149		USA
Goldhaber and Brewer (2000)	PD	1988, 1990, 1992	Level, pre	12	3786	2098	USA
Hanushek (1992)	EX2	1971–1975	Gains	2–6	441	22	USA
Harris and Sass (2009)	PD	2000–2003	Gains	3–10	+75k	32,000	USA
Harris and Sass (2011)	PD	1999–2004	Gains	3–10	±250k	1300	USA
Hill et al. (2005)	PD	2000–2003	Gain, pre	1&3	2963	699	USA
Holmlund and Sund (2008)	PD	1997–2004	Level	10–12	16,200		Sweden
Jepsen (2005)	PD	1991–1994	Gains	1&3	2652		USA
Kane et al. (2008)	PD	1999–2004	Gains	4–8	±300k		USA
Krieg (2005)	PD	2002–2003	Level, pre	4	49,415		USA
Kukla-Acevedo (2009)	PD	2000–2002	Gains	5	3812	120	USA
Metzler and Woessmann (2012)	PD	2004	WSBS	6	4302		Peru
Monk (1994)	PD	1987–1990	Level, pre	10–11	2829	1091	USA
Mullens et al. (1996)	PD	1990–1991	Gains	3	1043		Belize

(*Continued*)

Journal of Economic Surveys (2018) Vol. 32, No. 3, pp. 848–877

Table 1. *Continued*

Study	Method	Observation years	Measure	Grades	Students	Teachers	Country
Muñoz and Chang (2007)	PD	2005–2006	Gains	9	1487	58	USA
Muralidharan and Sheth (2016)	PD	2005–2009	Level, pre	1–5	66,660		India
Murnane and Phillips (1981)	PD	1973–1975	Level, pre	3–6	814		USA
Neugebauer et al. (2011)	PD	2001	Level	4	5858		Germany
Neild et al. (2009)	PD	2003	Level, pre	5–8	22,853	539	USA
Nye et al. (2004)	EX1	1985–1989	Both	K-3	6377		USA
Palardy and Rumberger (2008)	PD	1999	Level, pre	1	3496		USA
Papay and Kraft (2015)	PD	2001 and 2009	Within-teacher	4–8	100k+	±9000	USA
Rockoff (2004)	PD	1989–2000	Level	K-6	±10k	297	USA
Rowan et al. (1997)	PD	1988	Level, pre	10	5381		USA
Sharkey and Goldhaber (2008)	PD	1990 and 1992	Gains	12	486	224	USA
Sokal et al. (2007)	EX2	2006?	Gains	3–4	175	RA	Canada
Summers and Wolfe (1977)	PD	1968–1971	Gains	6	627		USA
Winters et al. (2013)	PD	2000–2004	Both	3–10	±1700k	±13k	USA

PD = Panel data study; EX1 = Experimental study with both teachers and students randomly assigned; EX2 = Experimental study with students randomly assigned; Level, pre = Test scores in levels with control for pretest score; Both = Level, pre and gains; WSBS = Within Student Between Subject; M = Matching.

Table 2. Description of Studies Included in Literature Review: Not Published in Peer-Reviewed Journals

Study	Method	Observation years	Measure	Grades	Students	Teachers	Country
Ammermüller and Dolton (2006)	PD	1995, 1999, 2001, 2003	Gains	4,8	±8k		USA and UK
Antecol et al. (2014)	EX2	2001–2002	Level	1–5	1900	100	USA
Coenen et al. (2014)	PD	2010–2012	Both	3–5	3602	202	The Netherlands
Betts et al. (2003)	PD	1997–1999	Gains	2–9?			USA
Cantrell et al. (2008)	EX2	2003–2004	Both	2–5	3790	99	USA
Cavalluzzo (2004)	PD	2000–2002	Level, pre	9–10	103k	2109	USA
Cavalluzzo et al. (2015)	PD	2008–2012	Level, pre	8–11	±250k	13,284	USA
Constantine et al. (2009)	EX2	2003–2005	Level	K-5	2600	174	USA
Cowan and Goldhaber (2015)	PD	2006–2013	Gains	4–8	±1300k	+/− 12k	USA
Decker et al. (2004)	EX2	2001–2002	Level	1–5	±2000		USA
Goldhaber and Brewer (1996)	PD	1990	Level, pre	10	6196		USA
Ladd and Sorensen (2015a)	PD	2006–2011	Level, pre	6–12	±2500k		USA
Ladd and Sorensen (2015b)	PD	2006–2011	Level, pre	6–8	±1300k		USA
Sanders et al. (2005)	PD	1999–2002	Both	4–8	37k	122	USA

PD = Panel data study; EX1 = Experimental study with both teachers and students randomly assigned; EX2 = Experimental study with students randomly assigned; Level, pre = Test scores in levels with control for pretest score; Both = Level, pre and gains; WSBS = Within Student Between Subject; M = Matching.

design (EX1 or EX2). Of these experimental studies, all but two have students randomly assigned to classes, but not teachers (EX2). Included studies generally examine the effect of teacher characteristics on student test scores by considering achievement levels (Level), achievement levels with controls for prior achievement (Level, pre) or achievement gains between two consecutive school years (Gains). One study considered within-student between-subject test score differences (WSBS), and one study evaluated achievement levels using statistical matching to control for prior achievement differences (Level, M).

3. Acquired Teacher Characteristics and Student Test Scores

This section outlines the literature on how *acquired* teacher characteristics affect student test scores. The following four subsections are distinguished:

(a) Advanced degrees and college quality
(b) Teaching certificates
(c) Test scores
(d) Experience

The first category focuses on the extent to which teacher degrees, and the quality of the undergraduate college attended, matter for student achievement. The second category focuses on results for different teaching certificates. These studies generally focus on subject-specific certifications and on (alternative) certification pathways, such as Teach For America (TFA) and National Board Certified Teachers (NBCT). The third category focuses teacher test scores, often used as a proxy for teacher ability. The fourth category evaluates the extensive literature on the effects of teacher experience. The results for sociodemographic teacher characteristics (i.e. gender and race) are covered in section four.

3.1 *Advanced Degrees and College Quality*

It may seem straightforward to expect a positive relationship between teacher education level and student learning. One might argue, for example, that teacher education level is a reasonable proxy for teacher quality. This result would be concurrent with both human capital and signalling theory. On the other hand, it may be that being higher educated does not automatically translate into higher student test scores after a certain education threshold level has been obtained. Alternatively, evaluating generic education levels does not acknowledge that there might be important differences in college quality. Most empirical work merely focuses on the generic effect of having a Master's or PhD degree, relative to a Bachelor's degree. The relationship between teacher education level and student test scores found by these studies is thus expected to be either positive or nonsignificant. Other empirical studies do focus on the actual quality of the colleges teachers attended, or on the coursework taken.

Table 3 presents the main findings for highest attained education level and subject-specific degrees. The second column indicates if the study considers primary or secondary education and the third column whether the evidence was provided by an experimental study (EXP) or panel data (type I, II or III).[2] The table is sorted based on the information in this column as it is informative for the quality of evidence that is delivered by the study. The next level of sorting is based on how recent the studies are. The fourth column indicates the teacher education level(s) the findings relate to, whereas the fifth column which subjects are considered. Finally, the last two columns present the nature of the effect that was found and, if mentioned in the study, the reported effect size. In some cases a range is provided, or multiple estimates.

The empirical results suggest that there is no positive significant relationship between the teacher education level and student test scores. This empirical finding appears to be robust as this absence of a relationship is found by both older and more recent studies, across different grade levels (see Tables 1 and 2) and for different subjects. Some studies, particularly those in which students are not randomly assigned

Table 3. Findings of Studies on Teacher Education Level

Study	Level	Type	Focus	Subjects	Results	
					Relation	Effect
Published						
Hanushek (1992)	PE	I	Master	Reading/vocabulary	0/0	
Murnane and Phillips (1981)	PE	I	Master	Vocabulary	0	
Harris and Sass (2011)	PE/SE	II	AD newly acquired	Math/reading	El.: 0/0 Mi.: +/−	
Clotfelter et al. (2010)	SE	II	AD	WSBS: Math/English/biology/ELP	0	
Dee and Cohodes (2008)	SE	II	GD	WSBS: Math/reading/science/social studies	0	
Aaronson et al. (2007)	SE	II	B major, M, PhD	Math	0/0/0	
Clotfelter et al. (2007)	PE	II	M, AD, PhD	Math/reading	M: 0/− A: −/− PhD: −/0	M: −0.004 to 0.008 A: −0.025 to 0.052 PhD: −0.093
Croninger et al. (2007)	PE	II	AD, Elementary education degree	Math/reading	0/0, 0/+	0.078
Clotfelter et al. (2006)	PE	II	AD	Math/reading	−/−	−0.028/−0.023
Jepsen (2005)	PE	II	B, B+ (ref. M or more)	Math/reading	0/0, 0/0, 0/0	
Muñoz and Chang (2007)	SE	III	Master or higher	Reading	0	
Goldhaber and Brewer (2000)	SE	III	M, PhD, B/M major S, B major E, M major E	Math/science	0/0, 0/0, +/0, −/0, 0/0	

(*Continued*)

Journal of Economic Surveys (2018) Vol. 32, No. 3, pp. 848–877

Table 3. *Continued*

					Results	
Study	Level	Type	Focus	Subjects	Relation	Effect
Rowan et al. (1997)	SE	III	B or M major in math	Math	+	0.016
Goldhaber and Brewer (1997)	SE	III	B, M major in math	Math	+, +	
Monk (1994)	SE	III	M, M+	Math/science	G10: −/−, 0/−G11: 0/0, 0/−	
Not published						
Ladd and Sorensen (2015a)	PE/SE	II	Master	Math/reading (Mi), 8 subjects (Hi)	0/0, 0/0/0/0/0/0/0/0	
Betts et al. (2003)	PE/SE	III	Master, PhD (Mi and Hi only)	Math/reading	E: +/0 M: 0/0, 0/0 H: 0/+, 0/+	
Goldhaber and Brewer (1996)	SE	III	M, B in subject, M in subject	Math/science/English/history	0/0/0/0, +/+/0/0, +/0/0/0	

Notes: Effect sizes are given when interpretable, available in the studies and statistically significant. Different estimated coefficients are separated by commas; estimations of the same variables for different subjects are separated by a slash sign (/);for studies with multiple estimations, ranges are provided. PE = Primary Education; SE = Secondary Education; B = Bachelor, M = Master, B+ = Bachelor degree plus additional credentials. M+ = Master degree plus additional credentials. AD = Advanced degree (higher than bachelor), GD = Graduate Degree, Major S = Major in subject, Major E = Major in education, El = Elementary school, Mi = Middle school, Hi = High school; WSBS = within-student between-subject model. No separate estimations for all subjects but one estimation using WITHIN-STUDENT BETWEEN-SUBJECT effects. G10/11: Grade 10/11.

to teachers, even report a negative association between teacher education level and student performance. Theoretically, it is unlikely that there indeed such a negative effect and these findings may thus reflect that teachers are selectively assigned to classes.

Three published studies focus on the effects of subject-specific Bachelor's and Master's degrees with majors in math or science. These all find a positive association between subject-specific education and student math test scores (i.e. Goldhaber and Brewer, 1997, 2000; Rowan et al., 1997). An unpublished study by Goldhaber and Brewer (1996) also finds positive effects for math and science, but not for English and History. In general, the best evidence available to date suggests that subject-specific bachelor and master degrees in math or science are positively related to student test scores. A cautionary note, however, is that these results are all from panel data studies without randomizing students and teachers to classes (type III), as to explicitly account for unobserved heterogeneity.

Two results merit additional discussion. Harris and Sass (2011) find that students achieve better math test scores, but lower reading test scores if teachers have *recently* acquired an advanced degree. This might seem contradictory, but the study interestingly suggests that the timing of acquiring an advanced degree can determine the nature of the effect it has on student test scores. A recent unpublished study by Ladd and Sorensen (2015a) find no significant effect of having a Master's degree on student test scores. A relative unique feature of their research design is that it includes teacher fixed effects, thereby taking into account that teachers choose whether or not to obtain a degree and at what time in their career.

Table 4 summarizes the findings for the quality of (undergraduate) college attended. The studies included all use Barron's ranking of college selectivity, thereby distinguishing four categories: very competitive, competitive, not competitive and unranked. Three of these five studies find that students of teachers who attended more competitive colleges perform better. Boyd et al. (2008) and Clotfelter et al. (2010) report small effect sizes which are remarkably similar (0.014 and 0.019 of a standard deviation, respectively). Clotfelter et al. (2010) report a statistically significant and positive effect for Grade 10 students, but find no effect for Grade 5 students. This may suggest that college quality is important for student test scores in secondary education only. Positive effects are found for math (Boyd et al., 2008) and for a general test (Summers and Wolfe, 1977), whereas none of the studies find positive effects for reading scores.

3.2 *Teaching Certificates*

A substantial amount of teachers in the USA are not officially certified to teach. This helps explain why the included empirical studies on this topic all focus on the USA. The available empirical literature on teacher certification can broadly be divided into three categories: (a) studies evaluating the effects of traditional certification versus alternative pathways, (b) studies evaluating the effects of teacher certification in particular subjects, and (c) studies evaluating if students of National Board Certified teachers achieve higher test scores.

Table 5 presents the results for the effects of traditional versus alternative pathways on student test scores. Note that most studies to some extent differ in the focus they adopt. Some compare several particular alternative pathways to traditional certification, whereas others do not distinguish between the various types of alternative certification.

In most studies, no significant differences in student performance is found, although some find negative outcomes for alternative pathways. However, given that none of the results are based on experimental research designs, this might reflect other unobserved differences between traditionally and alternatively certified teachers. For example, when evaluating TFA, taking such selection effects into account is very important, as TFA teachers tend to have very different background characteristics compared to other teachers. Another potential issue is that alternatively certified teachers can still obtain regular certification later on in their careers. Being traditionally certified thus correlates with other characteristics that may

influence student test scores, such as experience. For example, TFA, attracts mainly uncertified young, and thus relatively inexperienced, teachers. The negative result for TFA in Kane et al. (2008) may thus partially reflect an experience effect, rather than the effect of TFA certification.

Kane et al. (2008) evaluate regular certified teachers, regular uncertified teachers and three types of alternatively certified teachers in New York City. Given the small estimation coefficients, the authors conclude that a teacher's classroom performance during the first 2 years of teaching is a far more reliable indicator of future effectiveness.

Boyd et al. (2006) also compare different pathways into teaching in New York City. They investigate whether teachers who enter through alternative routes into teaching with reduced coursework are as effective as other teachers. Their results reveal that teachers licensed through these new programs attain higher achievement gains than temporary license teachers. Yet, when compared to university trained teachers, these teachers generally realize smaller achievement gains (from 2% to 5% of a standard deviation) in both mathematics and English language arts. Most of these differences, though, disappear over time as the teacher cohort matures. In this respect, Boyd et al. (2008) find very similar results.

The two unpublished studies on this topic do not display very different results. However, Decker et al. (2004), using a research design with random assignment of students to teachers (type 1) find that TFA-certified teachers actually have a positive effect on student performance.

Table 4. Findings of Studies on Undergraduate College Quality of Teachers

					Results	
Study	Level	Type	Focus	Subjects	Relation	Effect
Published						
Clotfelter et al. (2010)	SE	II	Very competitive, competitive, unranked (ref. not competitive)	WSBS: Math/English/ biology/ELP	+, 0, 0	0.019
Boyd et al. (2008)	PE	II	Most competitive, competitive, least competitive (ref. less competitive)	Math	0, +, 0	0.014
Clotfelter et al. (2006)	PE	II	Very competitive, competitive, unranked (ref. less competitive)	Math/reading	0/0, 0/0, 0/0	
Murnane and Phillips (1981)	PE	III	Dummy attended prestigious college	Vocabulary	0	
Summers and Wolfe (1977)	PE	III	Rating of teacher's college	General	+	

Notes: Effect sizes are given when interpretable, available in the studies and statistically significant. Different estimated coefficients are separated by commas; estimations of the same variables for different subjects are separated by a slash sign (/); for studies with multiple estimations, ranges are provided. PE = Primary Education; SE = Secondary Education; CQ = College Quality; WSBS = within-student between-subject model. No separate estimations for all subjects but one estimation using within-student between-subject effects.

Table 5. Findings on Teacher Certification: Alternative Pathways into Teaching

Study	Level	Type	Focus	Subjects	Results: Relation	Results: Effect
Published						
Kane et al. (2008)	PE/SE	II	Fellows, TFA, international recruits, other uncertified (ref: regular)	Math/reading	0/−, +/0, −/0, 0/0	−0.012, 0.018, −0.027
Boyd et al. (2008)	PE	II	College recommended, fellows, TFA, IE, other	Math	+,0,0,0,0	0.031
Sharkey and Goldhaber (2008)	SE	II	Temp or emerg. certification, private school certification	Math/science	+/0, 0/0	
Croninger et al. (2007)	PE	II	Regular or alt. (ref: none, temp, prov, emerg, prob	Math/reading	0/0	
Boyd et al. (2006)	PE/SE	II	IE, fellows, TFA, Temp, other	Math/ELA	−/0, −/−, 0/−, −/−, −/−	−0.012 to −0.031
Jepsen (2005)	PE	II	Not fully certified (ref. fully certified)	Math/reading	0/0	
Palardy and Rumberger (2008)	PE	III	Full certification (ref. no full certification)	Math/reading	+/+	
Darling-Hammond et al. (2005)	PE	III	Uncertified, alt., emerg or temp, out-of-field, no-test certified, certification code missing (ref. standard-certified)	Math/reading	−/−, −/0, −/0, −/+, −/−, −/−	
Darling-Hammond et al. (2005)	PE	III	TFA (ref. non-TFA)	Math/reading	−/−	
Goldhaber and Brewer (2000)	SE	III	Probationary, emergency in subject, private school certification	Math/science	0/0, 0/0, −/0	−0.010
Not published						
Constantine et al. (2009)	PE	I	Alt. (ref. regular)	Math/reading	0/0	
Decker et al. (2004)	PE	I	TFA	Math/reading	+/0	0.150−0.260

Notes: Effect sizes are given when interpretable, available in the studies and statistically significant. Different estimated coefficients are separated by commas; estimations of the same variables for different subjects are separated by a slash sign (/); for studies with multiple estimations, ranges are provided. PE = Primary Education; SE = Secondary Education; IE = Individual Evaluation; TFA = Teach For America; Alt. = Alternative; Temp, prov, emerg, prob = Temporary, provisional, emergency or probational certification.

Journal of Economic Surveys (2018) Vol. 32, No. 3, pp. 848–877

Table 6 summarizes the results of the seven studies examining effects of general and subject-specific teacher certification. Most of the analyses find no effects from certification in itself. Subject certification, on the other hand, is generally related to higher student test scores in these subjects, most notably mathematics (Dee and Cohodes, 2008; Neild et al., 2009; Clotfelter et al., 2010). Such a clear relationship is not found for all subjects, though. Clotfelter et al. (2010) use data on high school students in Grade 9 and Grade 10. On average, they find that students of teachers with certification in a subject or in a related subject area perform respectively 0.070 and 0.051 of a standard deviation better. Yet, when disaggregated by subject, the most positive results are found for mathematics (0.11 of a standard deviation) and English (0.10 of a standard deviation).

Dee and Cohodes (2008) find a similar coefficient for students from math-certified teachers. These students perform 0.12 of a standard deviation better than teachers without certification in math. In addition, the authors also find a positive effect of 0.08 of a standard deviation for subject-certified teachers in social sciences. For subject-certified teachers in reading and science, no differences are found.

Boyd et al. (2008) find negative effects on math performance for noncertified teachers, but not for other subjects. According to their results, students from teachers without any certification perform 0.04 of a standard deviation lower in math. However, another recent study finds no significant relationship between subject-certified teachers and student test scores in math and science (Sharkey and Goldhaber, 2008).

Table 7 summarizes the main findings of the 10 studies examining National Board Certification (NBC). Despite relatively solid research designs, five of these are not (yet) published. While some studies only compare student outcomes for NBC teachers with other teachers, other studies also investigate whether NBC effects can be explained by signalling or screening, or also by human capital effects resulting from the application process.

From the five published studies, four find that students from National Board Certified Teachers perform better (Clotfelter et al., 2006, 2007, 2010; Goldhaber and Anthony, 2007), while only one finds no differences between teachers who are certified, unsuccessful applicants or nonapplicants (Harris and Sass, 2009). In three consecutive studies, Clotfelter et al. (2006, 2007, 2010) find mainly positive effects of having a teacher who is national board certified, with the exception of math test scores for Grade 5 students (Clotfelter et al., 2006). Results in Goldhaber and Anthony (2007) also suggest positive outcomes for students with a NBC teacher. In addition, they also distinguish between current and future certification. In this way, they are able to examine whether National Board Certification is merely a signal of high-quality teachers, or if the process of obtaining the certification has a quality improvement effect on teachers as well. For math, they find evidence for both. Students of teachers who would receive certification later on perform 0.05 of a standard deviation better in math, and another 0.04 of a standard deviation better if their teacher already was National Board Certified. For reading, smaller effects are found.

Nonpublished findings regarding NBC certification are mostly in line with the published results summarized above. Two recent studies find positive effects on student performance in both math and reading (Cavalluzzo et al., 2015; Cowan and Goldhaber, 2015). Cavalluzzo (2004) found similar results and Cantrell et al. (2008), comparing both successful applicants and nonapplicants to failed applicants also show that the latter group performs worse in terms of student performance. One unpublished study, by Sanders et al. (2005), did not find any effect of NBC teachers overall.

In a recent report, aiming to explain this body of results, Cavalluzzo et al. (2015) conclude that there are signalling and screening effects of National Board Certification, but no actual human capital effect. Apparently, National Board Certification can distinguish more effective from less effective teachers, but the actual process of becoming NBC-certified does not make teachers more effective. Furthermore, Cowan and Goldhaber (2015), based on a value-added model, also conclude that teachers with National Board Certification are more effective, both in math and reading and both in elementary and middle school, although effect sizes vary.

Table 6. Findings on Teacher Certification

Study	Level	Type	Focus	Subjects	Results: Relation	Results: Effect
Published						
Clotfelter et al. (2010)	SE	II	Certified in subject, in related subject (ref. no certification)	WSBS: Math/ English/biology/ELP	+, +	0.070/0.051
Boyd et al. (2008)	PE	II	Not certified, in Math, Science, Special ed., other (ref. certified)	Math	−, 0, 0, 0, 0	−0.042
Dee and Cohodes (2008)	SE	II	Subject	Math/reading/science/social studies	+/0/0/+	0.116/0.081
Sharkey and Goldhaber (2008)	SE	II	Not subject-certified	Math/science	0/0	
Neild et al. (2009)	PE/SE	III	Secondary subject-certified, special ed, other field, not certified (ref. primary subject-certified)	Math/science	0/+, −/−0/−, −/−	
Goldhaber and Brewer (2000)	SE	III	Prob cert. in subject, emerg cert. in subject, private school cert., not certified in subject (ref. standard cert. in subject	Math/science	0/−, 0/0, −/0, −/0	
Goldhaber and Brewer (1997)	SE	III	Certified, in Math	Math	−, +	

Notes: Effect sizes are given when interpretable, available in the studies and statistically significant. Different estimated coefficients are separated by commas; estimations of the same variables for different subjects are separated by a slash sign (/); for studies with multiple estimations, ranges are provided. PE = primary education; SE = secondary education; WSBS = within-student between-subject model. No separate estimations for all subjects but one estimation using within-student between-subject effects.

Table 7. Findings on Teacher Certification: NBC

Study	Level	Type	Focus	Subjects	Results	
					Relation	Effect
Published						
Clotfelter et al. (2010)	SE	II	Precertification, application, NBC	WSBS: Math/English/biology/ELP	+, +, +	0.022/0.049/0.049
Harris and Sass (2009)	PE/SE	II	NBC	Math/reading	0/0	
Clotfelter et al. (2007)	PE	II	NBC	Math/reading	+/+	0.018 to 0.061/0.012 to 0.038
Clotfelter et al. (2006)	PE	II	NBC	Math/reading	0/+	0.030
Goldhaber and Anthony (2007)	PE	III	Future NBC, current NBC (ref. No NBC)	Math/reading	+/+, 0/+	0.050/0.040, 0.020
Not published						
Cantrell et al. (2008)	PE	I	Nonapplicants, unsuccessful applicants (ref. NBC)	Math/language	0/0, −/−	−0.173/−0.134
Cavalluzzo et al. (2015)	SE	II	NBC (ref. no NBC), ever NBC (ref. never), past, current (ref. future applicant)	Math/English/science	+/+/0, +/+/0, 0/0/0, 0/0/0	0.070 to 0.078/0.026 to 0.062, 0.036 to 0.071/0.056
Cowan and Goldhaber (2015)	PE/SE	II	NBC (ref. no NBC)	Math/reading	+/+	0.017 to 0.051/0.011 to 0.033
Sanders et al. (2005)	PE/SE	II	NBC (ref. others)	Math/reading	0/0	
Cavalluzzo (2004)	SE	II	NBC, application, failed/withdrawn (ref. No NBC)	Math	+, +, −	0.074/0.019/−0.026

Notes: Effect sizes are given when interpretable, available in the studies and statistically significant. Different estimated coefficients are separated by commas; estimations of the same variables for different subjects are separated by a slash sign (/); for studies with multiple estimations, ranges are provided. PE = primary education; SE = secondary education; NBC = National Board Certification. WSBS = within-student between-subject model. No separate estimations for all subjects but one estimation using within-student between-subject effects.

Journal of Economic Surveys (2018) Vol. 32, No. 3, pp. 848–877

Highlighting the importance of selection effects, Cantrell et al. (2008) conducted an experiment in which they randomly assign elementary students in the Los Angeles region to either NBC applicants (successful or unsuccessful) or to comparison teachers in the same school. They find that certified teachers were not more effective than nonapplicants. Furthermore, teachers who unsuccessfully applied were less effective, compared to both certified teachers and nonapplicants. Students from unsuccessful applicants performed 0.17 of a standard deviation worse in math and 0.13 of a standard deviation worse in language, compared to students from certified or nonapplicant teachers. Their findings do suggest that the National Board for Professional Teaching Standards (NBPTS) is able to distinguish between high- and low-quality teachers. However, when certified teachers were compared to the large majority of teachers who never applied for NBC, no significant differences in student test scores are found.

Similarly, Cavalluzzo (2004) did find differential effects for different subgroups of teachers. Students from certified teachers performed 0.07 of a standard deviation better in math than students from noncertified teachers (who did not apply). Students from teachers who were still in the application process also performed slightly better, but only by 0.02 of a standard deviation. Students from failed or withdrawn teachers performed 0.03 of a standard deviation worse than students from noncertified teachers who never applied.

3.3 *Teacher Test Scores*

The empirical literature on the effects of teacher quality on student test scores generally departs from recognizing that teacher ability cannot be directly observed, but that it can be approximated teachers own test scores. Test scores used in this respect are licensure test scores, test scores on subject knowledge, verbal skills tests and GPA or SAT scores.

Two main conclusions can be drawn from the empirical findings in Table 8. The first is that overall test scores (i.e. GPA and SAT scores) and math test scores both appear good approximations of teacher ability, but only for math-related subjects (see, among others, Clotfelter et al., 2006, 2007; Boyd et al., 2008). The work by Metzler and Woessman (2012) is particularly interesting, because they compare teacher and student test scores on similar (but not identical) tests. Their findings also indicate that these scores correlate for math, but not for reading. This implies that that student math performance is higher for teacher with greater subject-specific knowledge.

Coenen et al. (2014) support this first main finding, but instead of using teacher test score data as a proxy for teacher ability, they use information on how teachers were tracked in secondary education in the Netherlands. This tracking decision is a direct consequence of the final standardized and nationwide tests at the end of primary education. The results indicate that students from teachers who were in the lowest secondary education track (i.e. lower test scores at age 12) performed 0.20 of a standard deviation lower in math scores, compared to students from teachers who were in the intermediate or highest secondary education track. For reading, no significant effects are found.

The latter result brings us to the second main finding: teachers' language-related test scores have no significant effect on student performance (see Murnane and Phillips, 1981; Hanushek, 1992). Clotfelter et al. (2010) also find a positive relationship between achieved licensure test scores of teachers and student performance in math and biology, but even a negative relationship is found with respect to students' English performance.

Kukla-Acevedo (2009) rightfully points out that student performance may be influenced jointly by both teacher qualifications (like GPA) and experience. Of all the indicators of teacher qualifications and experience examined, overall GPA is the only indicator which consistently is positively related to student test scores in math. However, this positive association is not constant over time, due to a combined effect of GPA and experience over time. In particular, teachers with a lower GPA appear to reduce the gap in teaching effectiveness once they become more experienced.

Table 8. Findings on Teacher Test Scores

Study	Level	Type	Type of test	Subjects	Results: Relation	Results: Effect
Published						
Hanushek (1992)	PE	I	Verbal skills	Reading/vocabulary	+/0	
Murnane and Phillips (1981)	PE	I	Verbal skills	Vocabulary	0	
Metzler and Woessmann (2012)	PE	II	Test scores similar tests teachers and students	Math/reading	+/0	0.087
Clotfelter et al. (2010)	SE	II	Licensure test: Praxis II	General/math/English/biology/ELP	+/+/−/+/0	0.007/0.047/ −0.022/0.016
Kukla-Acevedo (2009)	PE	II	Overall GPA, math GPA	Math	+,0	0.034−0.084
Boyd et al. (2008)	PE/SE	II	Math, verbal SAT	Math	+, −	0.041/−0.033
Clotfelter et al. (2007)	PE	II	Elementary/early childhood education test	Math/reading	+/+	0.015−0.068
Clotfelter et al. (2006)	PE	II	Licensure test	Math/reading	+/0	0.012
Hill et al. (2005)	PE	III	Subject knowledge test: math, reading	Math	+, 0	
Rowan et al. (1997)	SE	III	Math quiz test score	Math	+	0.020
Mullens et al. (1996)	PE	III	Math score end primary school	Math	+	
Summers and Wolfe (1977)	PE	III	National Teacher Exam Score	General	−	
Not published						
Coenen et al. (2014)	PE	II	Tracking based on test at age 12	Math/reading	+/0	0.198

Notes: Effect sizes are given when interpretable, available in the studies and statistically significant. Different estimated coefficients are separated by commas; estimations of the same variables for different subjects are separated by a slash sign (/); for studies with multiple estimations, ranges are provided. PE = primary education; SE = secondary education.

3.4 *Teacher Experience*

There is a vast body of literature examining the effect of teacher experience on student test scores. The results of the studies included in this review are summarized in Table 9. One important issue for this theme is which functional should be adopted when student test scores are regressed on teacher experience. Therefore, the functional form of experience is included in the fourth column of the table. Most studies assume a linear relationship between teacher experience and student test scores, whereas others use categorical variables (allowing for heterogeneous teacher experience effects across categories). Others distinguish only between the first years of experience and all the years following. The latter choice largely stems from earlier empirical findings suggesting that only the first years of experience make a difference in explaining variation in student test scores.

Only a few studies find that teacher experience does not matter at all (e.g. Goldhaber and Brewer, 1996, 1997; Aaronson et al., 2007; Muñoz and Chang, 2007).[3] Generally speaking, empirical studies published (roughly) before 2011 suggest that increasing teacher experience only initially contributes to student performance, but that after 3–5 years, additional experience is largely irrelevant with respect to explaining variation in student test scores (see, among others, Nye et al., 2004; Boyd et al., 2006; Boyd et al. 2008; Clotfelter et al., 2010). Interestingly, Boyd et al. (2006) recognized that relatively weak teachers are more likely to quit the teaching profession in the beginning of their careers. When controlling for teacher attrition, no teacher experience effects were found on student performance, highlighting the importance of potential selection effects.

In contrast, the findings in Clotfelter et al. (2006, 2007) actually suggest that there are teacher experience effects throughout the first 27 experience years. More recent evidence corroborates this result. Papay and Kraft (2015) find that there is an continuous beneficial effect of experience on student performance in math and reading throughout a teacher's career and well beyond the initial 3–5 years. This evidence is consistent with evidence provided by Clotfelter et al. (2006, 2007). Furthermore, the estimated effects appear to be robust against controlling for teacher attrition effects. Ladd and Sorensen (2015b) use teacher fixed effects, such that it is possible to disentangle experience effects from cohort effects. Their results also show large returns to experience for middle school teachers, not only in their first years, but also after many years of teaching.

The results in Carrell and West (2010) are worth mentioning separately, as their study focuses on experience effects in higher education instead. Using a design in which students are randomly assigned to either inexperienced lecturers or experienced associate or full professors, they find that students taught by less experienced teachers performed 0.70 of a standard deviation better in introductory courses. However, students taught by more experienced teachers performed 0.69 of a standard deviation better in more advanced courses. This result suggests that the nature and context of introductory and advanced courses is markedly different and that it may be optimal to selectively assign teachers to introductory and advanced courses based on their level of seniority/experience.

4. Sociodemographic Teacher Characteristics and Student Test Scores

This section outlines the literature on how *sociodemographic* teacher characteristics affect student test scores. The following two subsections are distinguished:

(a) Teacher gender
(b) Teacher race

Table 9. Findings on Teacher Experience

Study	Level	Type	Experience	Subjects	Results: Relation	Results: Effect
Published						
Nye et al. (2004)	PE	0	More than 3 years	Math/reading	K: n.a/0, G1: 0/0, G2: 0/+, G3: +/0	0.189/0.142
Carrell and West (2010)	HE	I	Novice vs experienced	Contemp. courses/follow on courses	−/+	−0.690/0.700
Hanushek (1992)	PE	I	Linear	Reading/vocabulary	+/+	
Murnane and Phillips (1981)	PE	I	<8, ≥8 < 15, ≥15 years	Vocabulary	+, −,+	
Papay and Kraft (2015)	PE/SE	II	Linear and categories	Math/reading	+/+	0.028 to 0.132/0.046 to 0.058
Harris and Sass (2011)	PE/SE	II	Categories: 1–2, 3–4, 5–9, 10–14, 15–24, 25+ (ref. 0)	Math/reading	El: +/+, +/+, 0/+, 0/+, 0/+, 0/+; Mi: +/+, +/0, +/+, +/+, +/+, +/0; Hi: 0/−, 0/−, −/−, −/−, −/−, −/−	0.040 to 0.160 −0.040 to −0.160
Clotfelter et al. (2010)	SE	II	1–2 years, 3–5 years, 4 more categories (ref. 0 years)	WSBS: Math/English/biology/ELP	+, +, 0	0.048 to 0.061
Kukla-Acevedo (2009)	PE	II	Interaction experience and teacher GPA	Math	+	
Boyd et al. (2008)	PE/SE	II	1st 3–5 years	Math	+	Max 0.060
Clotfelter et al. (2007)	PE	II	To 27 years, half in 1st 2	Math/reading	+/+	0.092/0.118 to 0.067/0.096
Aaronson et al. (2007)	SE	II	Linear	Math	0	

(*Continued*)

Table 9. *Continued*

Study	Level	Type	Experience	Subjects	Results: Relation	Results: Effect
Croninger et al. (2007)	PE	II	0–2 years, 5+ years (ref. 2006, 2008, 2014 years)	Math/reading	0/−, 0/0	0.055
Boyd et al. (2006)	PE/SE	II	2 years vs 1, 3 years vs 2, 4+ years vs 3	Math/ELA	+/+, +/+, 0/0	0.076/0.047, 0.030/0.018
Clotfelter et al. (2006)	PE	II	To 27 years, half in 1st 2	Math/reading	+/+	0.100/0.080
Jepsen (2005)	PE	II	Linear	Math/reading	1: 0/0 3: +/+	0.021/0.020
Rockoff (2004)	PE	II	1st 2 years, linear	Math/reading	+/0, +/+	0.1000.150/0.180
Muñoz and Chang (2007)	SE	III	Linear	Reading	0	
Goldhaber and Brewer (2000)	SE	III	Linear	Math/science	0/+	
Monk (1994)	SE	III	Linear	Math/science	10: 0/0 11: +/−	
Bosshardt and Watts (1990)	SE	III	Linear	Other/econ.	−/−	
Summers and Wolfe (1977)	PE	III	Linear	General	+/−	
Not published						
Ladd and Sorensen (2015b)	PE/SE	II	Categories	Math/ELA	+/+	0.070 to 0.185/0.018 to 0.104
Betts et al. (2003)	PE/SE	III	Novice vs experienced	Math/reading	0	
Goldhaber and Brewer (1996)	SE	III	Linear	Math/science/English/history	0/0/0/0	

Notes: Effect sizes are given when interpretable, available in the studies and statistically significant. Different estimated coefficients are separated by commas; estimations of the same variables for different subjects are separated by a slash sign (/); for studies with multiple estimations, ranges are provided. PE = primary education; SE = secondary education; HE = higher education. WSBS = within-student between-subject model. No separate estimations for all subjects but one estimation using within-student between-subject effects. K = Kindergarten; G1 = Grade 1; G2 = Grade 2; G3 = Grade 3.

4.1 *Teacher Gender*

Two mechanisms are generally mentioned through which student test scores can be influenced by teacher gender (Coenen and Van Klaveren, 2016). The first refers to the observed feminization of the teaching profession and states that this leads to a lack of male role models and a more feminine school climate. This mechanism hypothesizes that feminization of the teaching profession negatively influences test scores of male students and positively influences the test scores of female students. The second mechanism is gender bias which hypothesizes that teachers may favour (or disfavour) students from the same gender when grading. Generally, the empirical literature does not distinguish between both mechanisms, but rather departs from recognizing that one or both mechanisms can lead to a situation in which the combination of student and teacher gender influences student test scores (Coenen and Van Klaveren, 2016). However, the second mechanism can only affect grades and not standardized test scores. Therefore the included studies in this review are relevant with respect to the first mechanism only.

Table 10 summarizes the findings of 11 empirical studies on teacher gender effects. In contrast to, for example, teacher certification effects, the influence of teacher gender is examined for many different countries (i.e. for the USA, the Netherlands, Canada, Germany, Sweden and India). For Sweden, Canada, Germany and the Netherlands, no interaction effects between teacher and student gender were found (Sokal et al., 2007; Holmlund and Sund, 2008; Neugebauer et al., 2011; Coenen and Van Klaveren, 2016). For India a positive teacher gender effect is found (Muralidharan and Sheth, 2016), while for the USA, there are mixed results (Dee, 2007; Clotfelter et al., 2010; Antecol et al., 2014). There appears to be no relationship between the type of panel data used in a study and the nature of the empirical results obtained.

The results for European countries and Canada clearly indicate that the feminization of teaching in primary and secondary education does not seem to affect student test scores. As such, concerns expressed by policy makers and educators that this feminization has contributed to the gender gap between boys and girls, therefore seems unfounded.

The mixed findings of US studies are more difficult to interpret. Dee (2007) finds positive same-gender effects and argues that this does not provide explicit guidance as to the appropriate policy responses. Dee (2007) concludes that the main implication of this finding is: *'to underscore that the gender interactions between students and teachers do appear to constitute an important "environmental" influence of educational outcomes for both girls and boys (Dee,* 2007, *p.27)'.*

Antecol et al. (2014) do provide explicit policy advice as they focus on female teachers, while also considering a teacher's background in mathematics. Their findings suggest that there is actually no such thing as a same-gender effect for boys. However, female students only suffered from lower test scores if assigned to a female teacher without a strong math background. Instead, this negative effect actually becomes (marginally) positive if female students were assigned to a female teacher with a strong math background. This may suggest there may be a same-gender effect for female teachers, but that it is mediated by subject-related knowledge. This result supports our earlier conclusion that subject-related knowledge and Master's degrees, particularly for math and science, contribute to better student performance.

4.2 *Teacher Race*

The mechanisms through which student test scores can be influenced by teacher race are mostly similar to those mentioned in explaining teacher gender effects. The literature thus distinguishes between role model effects (i.e. teacher of the same race are supposedly better role models) and subjective grading effects (i.e. students may be differently assessed when judged by a same-race teacher). Since this review focuses on the effects of teacher characteristics on standardized student test scores, the empirical evidence on subjective grading effects is not considered here.

Table 11 summarizes the results of four published empirical studies on gender effects. Dee (2004) is the only study that focuses on primary education and uses Project STAR data in which teachers and students

Table 10. Findings on Gender Interactions

Study	Level	Type	Focus	Subjects	Results: Relation	Results: Effect
Published						
Sokal et al. (2007)	PE	I	Same-gender, boys	Reading	0	
Coenen and Van Klaveren (2016)	PE	II	Same-gender	Math	0	
Muralidharan and Sheth (2016)	PE	II	Same-gender	Math/language	+/+	0.032 to 0.036
Winters et al. (2013)	PE/SE	II	Teacher: Female	Math/reading	El: 0/0 Mi/Hi: +/+	0.013 to 0.028/0.007 to 0.016
Neugebauer et al. (2011)	PE	II	Same-gender	Math/German/science	0/0/0	
Clotfelter et al. (2010)	SE	II	Male T–female S, other comb.	WSBS: Math/English/biology/ELP	−, 0	−0.105
Holmlund and Sund (2008)	SE	II	Same-gender	Math/Swedish/English	0/0/0	
Dee (2007)	SE	II	Same-gender	WSBS: Math/science/English/history	+	0.045 to 0.050
Krieg (2005)	PE	III	Same-gender	Math/reading/writing/listening	0/0/0/0	
Ehrenberg et al. (1995)	SE	III	Gender	Math/science/English/history	0/0/0/0	
Not published						
Antecol et al. (2014)	PE	I	FT-FS, FT math major-FS, FT-MS	Math/reading	−/0, 0/0, 0/0	−0.080 to −0.200
Ammermüller and Dolton (2006)	PE/SE	II	Same-gender	Math/reading/science	Mixed	

Notes: Effect sizes are given when interpretable, available in the studies and statistically significant. Different estimated coefficients are separated by commas; estimations of the same variables for different subjects are separated by a slash sign (/); for studies with multiple estimations, ranges are provided. PE = primary education; SE = secondary education; E = elementary school; M = middle school; H = high school; FT = female teacher; MT = male teacher; FS = female student; MS = male student. WSBS = within-student between-subject model. No separate estimations for all subjects but one estimation using within-student between-subject effects.

Journal of Economic Surveys (2018) Vol. 32, No. 3, pp. 848–877

Table 11. Findings on Race Interactions

Study	Level	Type	Focus	Subjects	Results	
					Relation	Effect
Dee (2004)	PE	0	Same race	Math/reading	+/+	
Clotfelter et al. (2010)	SE	II	White T–Black S, White T–Hisp. S, Black T–White S	WSBS: Math/English/biology/ELP	+, +, −	0.036, 0.055, −0.083
Muñoz and Chang (2007)	SE	III	Race (minority)	Reading	0	
Ehrenberg et al. (1995)	SE	III	Race	Math/science/English/history	0/0/0/0	

Notes: Effect sizes are given when interpretable, available in the studies and statistically significant. Different estimated coefficients are separated by commas; estimations of the same variables for different subjects are separated by a slash sign (/); for studies with multiple estimations, ranges are provided. PE = primary education; SE = secondary education; T = teacher; S = student. WSBS = within-student between-subject model. No separate estimations for all subjects but one estimation using within-student between-subject effects.

Journal of Economic Surveys (2018) Vol. 32, No. 3, pp. 848–877

were randomly assigned to classrooms. Positive same-race effects are found for math and reading scores, but only for white and black teachers. No statistically significant effects were found for Asian or Hispanic teachers, but the authors interpret this as being mainly the result of relatively small sample sizes for these categories.

The other three studies focus on secondary education in the USA. Clotfelter et al. (2010) find that black and Hispanic students in classes with white teachers performed better (respectively 0.055 and 0.036 of a standard deviation). The other two studies by Ehrenberg et al. (1995) and Muñoz and Chang (2007) do not find any effects of same-race teacher–student interactions. Panel data studies that provide evidence for secondary education either consider the potential effects of unobserved heterogeneity (Clotfelter et al., 2010), or control for relevant background characteristics (Ehrenberg et al., 1995; Muñoz and Chang, 2007). The consequence of the absence of randomized assignment in these studies may be that the empirical outcomes are to a large extent driven by selection effects of teachers and students to schools. Clotfelter et al. (2010) indicate that there is actually an opposite-race effect, but their empirical results may also reflect that better neighbourhoods attract white teachers and, on average, better performing students. However, the same argument can be applied to the findings of Ehrenberg et al. (1995) and Muñoz and Chang (2007), in the sense that significant same-race effects are not found due to selection bias.

Considering both the limited number of available studies and the likelihood that selection effects may have biased most of the estimated effects, the main result is that the current evidence base is not unequivocal and strong enough to draw strong conclusions.

5. Conclusions and Discussion

Wayne and Youngs (2003) provide a solid overview of the empirical literature examining which teacher characteristics influence student test scores. A vast body of new empirical work has emerged over the past decade. Therefore, this study provides an up-to-date review. In addition, this review complements the existing knowledge base by extending the scope to results obtained outside the US context and to distinguish not only acquired teacher characteristics (e.g. experience, certification), but also sociodemographic teacher characteristics (i.e. gender and race).

The first main finding is that these more recent results corroborate Wayne and Youngs (2003) in that subject-related Master's degrees in math and science contribute to better student performance. Moreover, the empirical evidence indicates that overall test scores (i.e. GPA and SAT scores) and math performance of teachers are positively related to student's math performance. However, the same result is not found for subject-specific knowledge in language-related subjects.

Teacher certification, in general, has no positive effect on student performance, but subject-specific certification, especially in math, is frequently found to be positively related to student performance. Also, results for the National Board Certification reveal that the application process is able to both identify high- from low-quality teachers and to actually generate quality improvements. In general, alternative routes to teacher certification do not appear to harm student performance. When negative effects are reported, these often can be attributed to differences between teachers in experience and other unobserved characteristics. Furthermore, student performance differences between teachers from alternative paths are generally small (e.g. when compared to the predictive power of classroom performance in the first 2 years) and tend to disappear over time as the teacher cohort matures.

These abovementioned results for subject-specific knowledge and teacher certification (pathways) are important in formulating effective teacher certification policies. For example, in many countries such policies are still uniformly formulated with the objective to let teachers obtain an academic (Master's) degree, without explicitly acknowledging the potential importance of subject-specific knowledge. Another example is that teachers in primary education teach multiple subjects throughout the week, while perhaps student learning could be further improved if teachers specialize and teach only the subjects they master best. The relevance of subject-specific knowledge presented in this review highlights that there is a potential for more effective policies in this respect.

The results in this review also points to several new insights. First of all, the quality of the college a teacher has attended is relevant in explaining differences in student performance, yet only in secondary education. Future research could therefore explore whether it is possible to identify particular college characteristics (e.g. curriculum or learning materials) associated with college quality. These results would be informative both for policy makers and for teacher colleges currently not yet operating at that level of quality. A complicating factor is whether the college ratings used in the studies included in this review reflect actual differences in quality of the study programs, especially because the ratings are relative selectivity ratings, not absolute ratings of quality.

Until quite recently, the general consensus with respect to teacher experience has been that it contributes to student performance, but only in the first few years. After 3–5 years, the prevailing empirical evidence suggested that additional experience years did not seem to lead to additional learning gains. However, more recent studies, included in this review, provide new insights and conclude that experience continues to contribute to student test scores throughout a teacher's career. For future research, it therefore seems important not to focus primarily on the effectiveness of experience years, but rather on what the underlying mechanisms of such experience effects are. In other words, if teacher experience improves student performance, what important skills do more experienced teachers develop?

Since the review of Wayne and Youngs (2003), many more empirical studies have appeared that examine the effects of having a same-gender teacher. This surge in academic interest can be understood in light of the observed feminization of the teaching profession, but also because the gender achievement gap has changed, or even reversed, over time in favour of girls. Policy makers and educators alike are therefore worried that feminization of the teaching profession leads to a lack of male role-models, which would help explain why boys are relatively lagging behind in recent evaluations. However, the emerging body of empirical findings included in this review clearly show, at least for European countries and Canada, that feminization of the teaching profession does not differentially affect the test scores of boys and girls.

Finally, this review has also tried to summarize the empirical evidence on same-race teacher effects. However, considering both the limited number of available studies and the likelihood that selection effects may still have biased the estimated effects, we conclude that the evidence base is not yet unequivocal and strong enough to make any strong conclusions. More rigorous empirical work in this area is thus considered necessary.

Another important avenue for future research, in our opinion, would be to focus more on the mechanisms underlying the effects of particular teacher characteristics. As mentioned, many studies focus on the effects of teacher experience, whereas the more relevant question seems to be what important skills more experienced teachers develop. A better understanding of this would provide important input to improving teacher preparation programs. Similarly, empirical work would contribute greatly by examining, for each subject, which subject-specific knowledge teachers should possess. In general, empirical work on teacher effectiveness should move beyond evaluating characteristics and instead generate findings which can more directly be acted upon by policy makers and educators alike.

Acknowledgment

We would like to thank Marjolein Coonen for excellent research assistance.

1. Quasi-experimental evidence refers to studies using one of the following research designs: regression discontinuity, difference-in-difference analysis, Natural Experiments, IV-methods and statistical matching approaches.
2. We note that in this definition the US studies with data on Middle schools and/or High schools are both referred to as secondary education.
3. As is rightfully pointed out by Clotfelter et al. (2006), these studies will produce biased estimates in situations where non-random sorting of students and teachers into schools and classrooms introduce

correlations between the included characteristics and unobserved determinants of student test scores. Studies that try to address the issue of non-random sorting tend to find more positive estimates.

References

Aaronson, D., Barrow, L. and Sander, W. (2007) Teachers and student achievement in the Chicago public high schools. *Journal of Labor Economics* 25(1): 95–135.

Ammermüller, A. and Dolton, P.J. (2006) Pupil-teacher gender interaction effects on scholastic outcomes in England and the USA. Tech. rep., ZEW Discussion Papers.

Angrist, J.D. and Pischke, J.S. (2008) *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Antecol, H., Eren, O. and Ozbeklik, S. (2014). The effect of teacher gender on student achievement in primary school. *Journal of Labor Economics* 33(1), 63–89.

Betts, J.R., Zau, A. and Rice, L. (2003) *Determinants of Student Achievement: New Evidence from San Diego*. San Francisco, CA: Public Policy Institute of California.

Blundell, R. and Dias, M.C. (2009) Alternative approaches to evaluation in empirical microeconomics. *The Journal of Human Resources* 44(3): 565–640.

Bosshardt, W., and Watts, M. (1990) Instructor effects and their determinants in precollege economic education. *Journal of Economic Education* 21(3): 265–276.

Boyd, D., Grossman, P., Lankford, H., Loeb, S. and Wyckoff, J. (2006) How changes in entry requirements alter the teacher workforce and affect student achievement. *Education Finance and Policy* 1(2):176–216.

Boyd, D., Lankford, H., Loeb, S., Rockoff, J. and Wyckoff, J. (2008) The narrowing gap in New York City teacher qualifications and its implications for student achievement in high-poverty schools. *Journal of Policy Analysis and Management* 27(4): 793–818.

Cantrell, S., Fullerton, J., Kane, T.J. and Staiger, D.O. (2008) National board certification and teacher effectiveness: evidence from a random assignment experiment. Tech. rep., National Bureau of Economic Research.

Carrell, S.E. and West, J.E. (2010) Does professor quality matter? evidence from random assignment of students to professors. *Journal of Political Economy* 118(3): 409–432.

Cavalluzzo, L.C. (2004) *Is National Board Certification an Effective Signal of Teacher Quality*? Alexandria, VA: CNA Corporation.

Cavalluzzo, L., Barrow, L., Henderson, S., Mokher, C., Geraghty, T. and Sartain, L. (2015) *From large urban to small rural schools: an empirical study of National Board Certification and teaching effectiveness*. Final Report. Arlington, VA: CAN Corporation.

Chetty, R., Friedman, J.N. and Rockoff, J.E. (2014) Measuring the impacts of teachers I: evaluating bias in teacher value-added estimates. *American Economic Review* 104(9): 2593–2632.

Clotfelter, C.T., Ladd, H.F. and Vigdor, J.L. (2006) Teacher-student matching and the assessment of teacher effectiveness. *The Journal of Human Resources* 41(4): 778–820.

Clotfelter, C.T., Ladd, H.F. and Vigdor, J.L. (2007) Teacher credentials and student achievement: longitudinal analysis with student fixed effects. *Economics of Education Review* 26(6): 673–682.

Clotfelter, C.T., Ladd, H.F. and Vigdor, J.L. (2010) Teacher credentials and student achievement in high school a cross-subject analysis with student fixed effects. *Journal of Human Resources* 45(3): 655–681.

Coenen, J., Van Klaveren, C., Groot, W. and Maassen van den Brink, H. (2014) The effects of ability tracking of future primary school teachers on student performance, TIER Working Paper 14–7, 1–25.

Coenen, J. and Van Klaveren, C. (2016) Better test scores with a same-gender teacher. *European Sociological Review* 32(3), 452–464.

Constantine, J., Player, D., Silva, T., Hallgren, K., Grider, M. and Deke, J. (2009) *An Evaluation of Teachers Trained through Different Routes to Certification*. Jessup, MD: National Center for Education Evaluation and Regional Assistance.

Cowan, J. and Goldhaber, D.D. (2015) *National board certification and teacher effectiveness: evidence from Washington*. CEDR Working Paper 2015-3. Bothell, WA: University of Washington, Center for Education Data & Research.

Croninger, R.G., Rice, J.K., Rathbun, A. and Nishio, M. (2007) Teacher qualifications and early learning: effects of certification, degree, and experience on first-grade student achievement. *Economics of Education Review* 26(3): 312–324.

Darling-Hammond, L., Holtzman, D.J., Gatlin, S.J. and Heilig, J.V. (2005) Does teacher preparation matter? Evidence about teacher certification, Teach For America, and teacher effectiveness. *Education policy analysis archives* 13(42): 1–52.

Decker, P.T., Mayer, D.P. and Glazerman, S. (2004) *The Effects of Teach for America on Students: Findings from a National Evaluation*. Washington, DC: Mathematica.

Dee, T.S. (2004) Teachers, race, and student achievement in a randomized experiment. *The Review of Economics and Statistics* 86(1): 195–210.

Dee, T.S. (2007) Teachers and the gender gaps in student achievement. *The Journal of Human Resources* 42(3): 528–554.

Dee, T.S. and Cohodes, S.R. (2008) Out-of-field teachers and student achievement evidence from matched-pairs comparisons. *Public Finance Review* 36(1): 7–32.

Driessen, G. (2007) The feminization of primary education: effects of teachers sex on pupil achievement, attitudes and behaviour. *International Review of Education* 53(2): 183–203.

Ehrenberg, R.G., Brewer, D. and Goldhaber, D. (1995) Do teachers' race, gender, and ethnicity matter? Evidence from the NELS. *Industrial and Labor Relations Review* 48(3): 547–561.

Goldhaber, D.D. and Anthony, E. (2007) Can teacher quality be effectively assessed? National Board Certification as a signal of effective teaching. *The Review of Economics and Statistics* 89(1): 134–150.

Goldhaber, D.D. and Brewer, D.J. (1996) Evaluating the effect of teacher degree level on educational performance. In W.J. Fowler Jr (ed.), *Developments in school finance* (pp. 197–210). Washington, DC: National Center for Education Statistics. US Department of Education.

Goldhaber, D.D. and Brewer, D.J. (1997) Why don't schools and teachers seem to matter? Assessing the impact of unobservables on educational productivity. *The Journal of Human Resources* 32(3): 505–523.

Goldhaber, D.D. and Brewer, D.J. (2000) Does teacher certification matter? high school teacher certification status and student achievement. *Educational evaluation and policy analysis* 22(2): 129–145.

Hanushek, E.A. (1992) The trade-off between child quantity and quality. *Journal of Political Economy* 100(1): 84–117.

Hanushek, E.A. (2011) The economic value of higher teacher quality. *Economics of Education Review* 30(3): 466–479.

Harris, D. N. and Sass, T.R. (2009) The effects of NBPTS-certified teachers on student achievement. *Journal of Policy Analysis and Management* 28(1): 55–80.

Harris, D.N. and Sass, T.R. (2011) Teacher training, teacher quality and student achievement. *Journal of Public Economics* 95(7): 798–812.

Hill, H.C., Rowan, B. and Ball, D.L. (2005) Effects of teacher's mathematical knowledge for teaching on student achievement. *American Educational Research Journal* 42(2): 371–406.

Holmlund, H. and Sund, K. (2008) Is the gender gap in school performance affected by the sex of the teacher? *Labour Economics* 15(1): 37–53.

Jepsen, C. (2005) Teacher characteristics and student achievement: evidence from teacher surveys. *Journal of Urban Economics* 57(2): 302–319.

Kane, T.J., Rockoff, J.E. and Staiger, D.O. (2008) What does certification tell us about teacher effectiveness? Evidence from New York City. *Economics of Education Review* 27(6): 615–631.

Krieg, J.M. (2005) Student gender and teacher gender: what is the impact on high stakes test scores. *Current Issues in Education* 8(9).

Kukla-Acevedo, S. (2009) Do teacher characteristics matter? New results on the effects of teacher preparation on student achievement. *Economics of Education Review* 28(1): 49–57.

Ladd, H.F. and Sorensen, L.C. (2015a) Do Master's degrees matter? Advanced degrees, career paths and the effectiveness of teachers. CALDER Working Paper 136.

Ladd, H.F. and Sorensen, L.C. (2015b) Returns to teacher experience: student achievement and motivation in middle school. CALDER Working Paper 112.

Metzler, J. and Woessmann, L. (2012) The impact of teacher subject knowledge on student achievement: evidence from within-teacher within-student variation. *Journal of Development Economics* 99(2): 486–496.

Monk, D.H. (1994) Subject area preparation of secondary mathematics and science teachers and student achievement. *Economics of Education Review* 13(2): 125–145.
Mullens, J.E., Murnane, R.J. and Willett, J.B. (1996) The contribution of training and subject matter knowledge to teaching effectiveness: a multilevel analysis of longitudinal evidence from Belize. *Comparative Education Review* 40(2): 139–157.
Muñoz, M.A. and Chang, F.C. (2007) The elusive relationship between teacher characteristics and student academic growth: a longitudinal multilevel model for change. *Journal of Personnel Evaluation in Education* 20(3–4): 147–164.
Muralidharan, K. and Sheth, K. (2016) Bridging education gender gaps in developing countries: the role of female teachers. *Journal of Human Resources* 51(2): 269–297.
Murnane, R.J. and Phillips, B.R. (1981) What do effective teachers of inner-city children have in common? *Social Science Research* 10(1): 83–100.
Neild, R.C., Farley-Ripple, E.N. and Byrnes, V. (2009) The effect of teacher certification on middle grades achievement in an urban district. *Educational Policy* 23(5): 732–760.
Neugebauer, M., Helbig, M. and Landmann, A. (2011) Unmasking the myth of the same-sex teacher advantage. *European Sociological Review* 27(5): 669–689.
Nye, B., Konstantopoulos, S. and Hedges, L.V. (2004) How large are teacher effects? *Educational Evaluation and Policy Analysis* 26(3): 237–257.
Palardy, G.J. and Rumberger, R.W. (2008) Teacher effectiveness in first grade: the importance of background qualifications, attitudes, and instructional practices for student learning. *Educational Evaluation and Policy Analysis* 30(2): 111–140.
Papay, J.P. and Kraft, M.A. (2015) Productivity returns to experience in the teacher labor market: methodological challenges and new evidence on long-term career improvement. *Journal of Public Economics* 130: 105–119.
Rockoff, J.E. (2004) The impact of individual teachers on student achievement: evidence from panel data. *The American Economic Review* 94(2): 247–252.
Rothstein, J. (2009) Student sorting and bias in value-added estimation: selection on observables and unobservables. *Education Finance and Policy* 4(4): 537–571.
Rowan, B., Chiang, F.-S. and Miller, R.J. (1997) Using research on employees' performance to study the effects of teachers on students' achievement. *Sociology of Education* 70(4): 256–284.
Sanders, W.L., Ashton, J.J. and Wright, S.P. (2005) *Comparison of the effects of NBPTS certified teachers with other teachers on the rate of student academic progress*. Final report. Arlington, VA: National Board for Professional Teaching Standards.
Schwerdt, G. and Wuppermann, A.C. (2011) Is traditional teaching really all that bad? A within-student between-subject approach. *Economics of Education Review* 30(2): 365–379.
Sharkey, N.S. and Goldhaber, D. (2008) Teacher licensure status and student achievement: lessons from private schools. *Economics of Education Review* 27(5): 504–516.
Sokal, L., Katz, H., Chaszewski, L. and Wojcik, C. (2007) Good-bye, Mr. Chips: male teacher shortages and boys reading achievement. *Sex Roles* 56(9–10): 651–659.
Summers, A.A. and Wolfe, B.L. (1977) Do schools make a difference? *American Economic Review* 67(4): 639–652.
Van Klaveren, C. (2011) Lecturing style teaching and student performance. *Economics of Education Review* 30(4): 729–739.
Van Klaveren, C. and De Wolf, I. (2015) Systematic reviews in education research: when do effect studies provide evidence? In K. De Witte (ed.), *Contemporary Education Issues from an Economic Perspective* (pp. 1–26). Leuven: Leuven University Press.
Wayne, A.J. and Youngs, P. (2003) Teacher characteristics and student achievement gains: a review. *Review of Educational Research* 73(1): 89–122.
Winters, M.A., Haight, R.C., Swaim, T.T. and Pickering, K. (2013) The effect of same-gender teacher assignment on student achievement in the elementary and secondary grades: evidence from panel data. *Economics of Education Review* 34(C): 69–75.